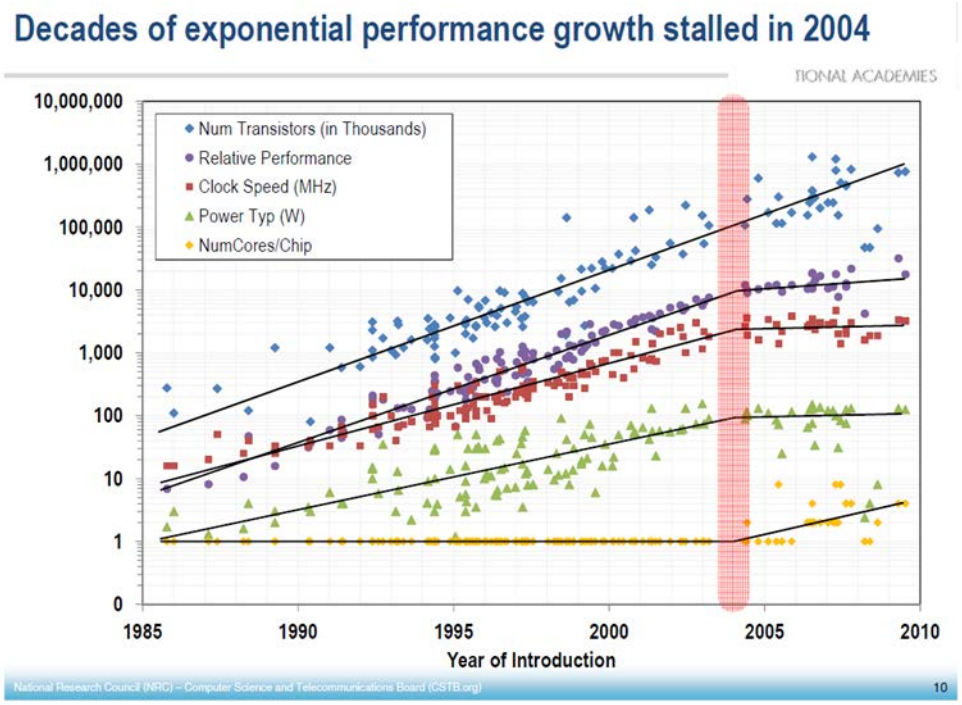
Power Efficiency Revolution For Embedded Computing Technologies (PERFECT)

DARPA MTO

Joe Cross – Program Manager

Introduction – where is processing headed

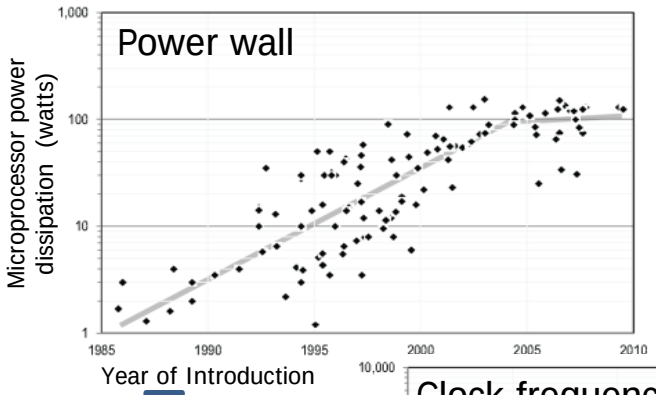


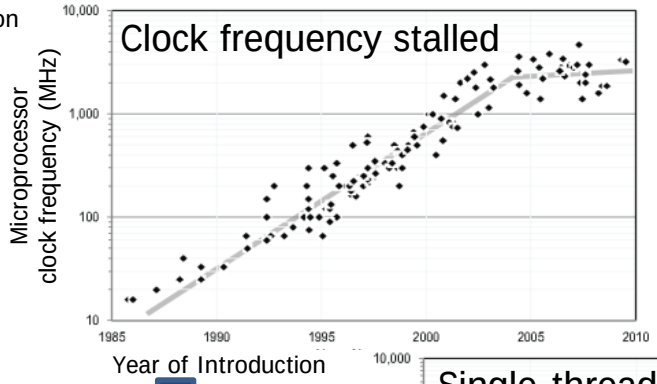
Source: NRC, The Future of Computing Performance, Game Over or Next Level?

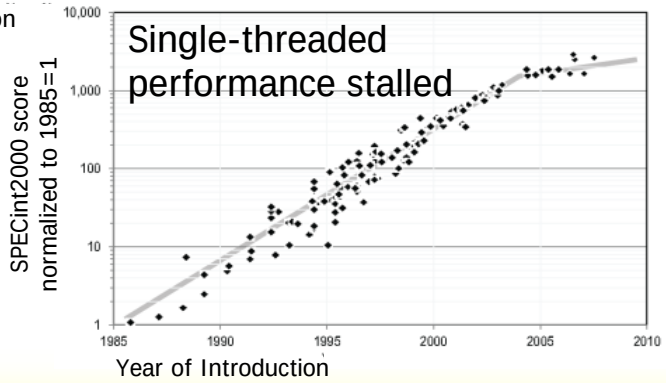
- Moore's law continues – we're getting more transistors with each geometry shrink.
- Dennard scaling has stopped – voltage decreases have stalled even as feature sizes shrink. Clock rates would have to decrease in order to hold power constant.
- Hardware offers lots more concurrency. Software in general can't use it all.

For power or energy constrained DoD embedded systems,
Greater power efficiency is the only path forward

Technology landscape: move past power limitations, effectively utilize concurrency

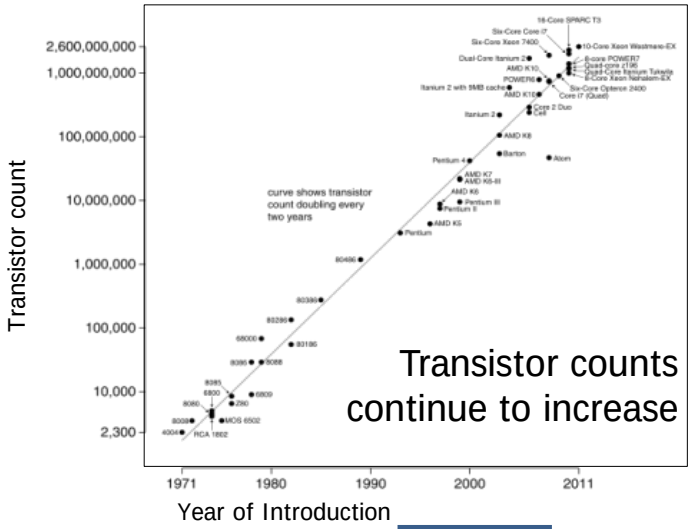






2011 NRC/CSTB Study: "The Future of Computing Performance"

Meanwhile:



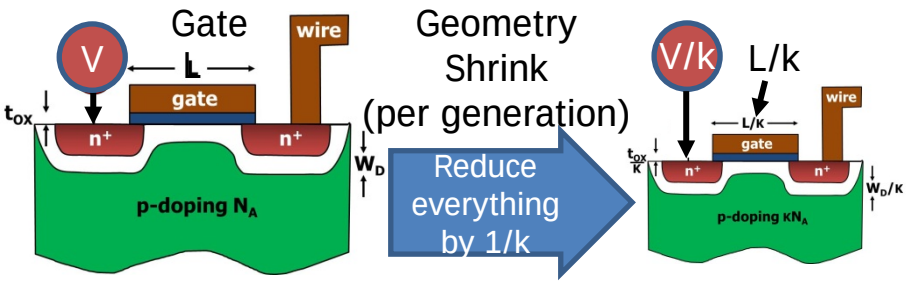
CONCURRENCY

- More but slower workers
- Potentially more performance
- Potentially more power efficiency

Concurrency only path left – National Academy of Science report

Industry's ride is over

The past: Dennard's Scaling



$$P_{density} = N_g C_{load} V^2 f$$

= power per unit area

N_g = CMOS gates/unit area

C_{load} = capacitive load/CMOS gate

V = supply voltage

f = clock frequency

k = scaling factor

k = typically 1.4 per geometry shrink

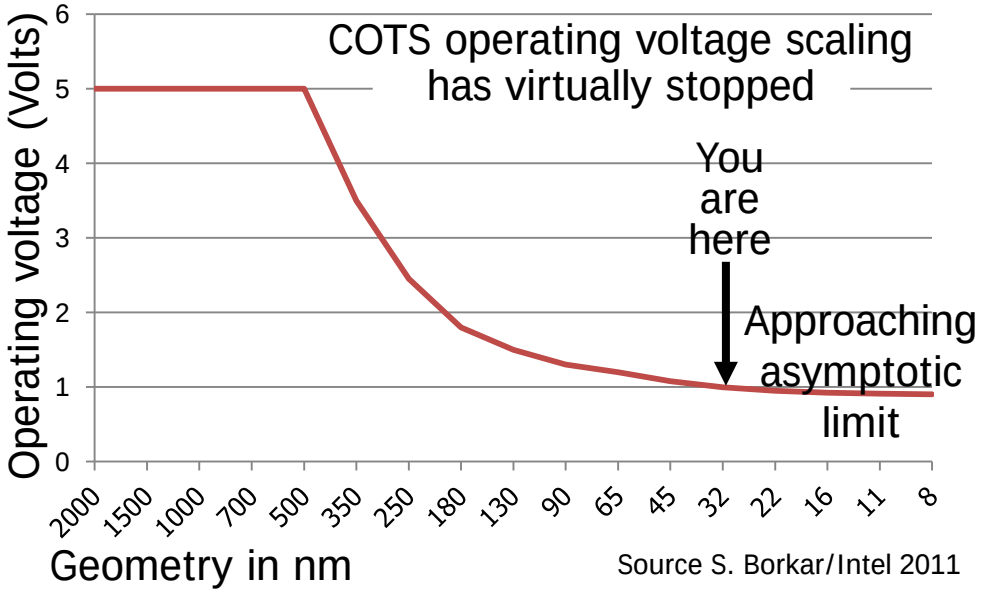
1/k = device feature scaling factor (typically 0.7 per geometry shrink)

For each generation/geometry shrink:

$$P_{density\ (scaling)} = (k^2)(1/k)(1/k^2)(k) = 1$$

Double the transistors (functionality) and increase the clock speed 40% per generation with the same power

Today: Dennard's Scaling is dead



Source S. Borkar/Intel 2011

$$P_{density\ (scaling)} = (k^2)(1/k)(1/k^2)(k) = k^2 \cong 2$$

But, power density cannot increase!

This physics is limiting COTS power efficiency to well below what we need for embedded sensor processing applications

If we could perform a factor of 50 more computational operations per watt[1], then we get processed real-time surveillance data directly from sensor platforms such as UAVs. No downlink limitations or lost data.

Other benefits would accrue, although not as well quantified:

- Better autonomous operation of air, ground, and undersea vehicles
- Better integration of data from sensor systems into actionable information

Better power efficiency is not a new goal. The approach recommended is to operate near threshold voltage. This leads to a set of interconnected enabling research areas:

- Resilience
- Parallelism
- Non-recurring cost, both hardware and software
- Resilience
- Verifiability
- Portability

[1]Substitute energy for power if you're battery operated

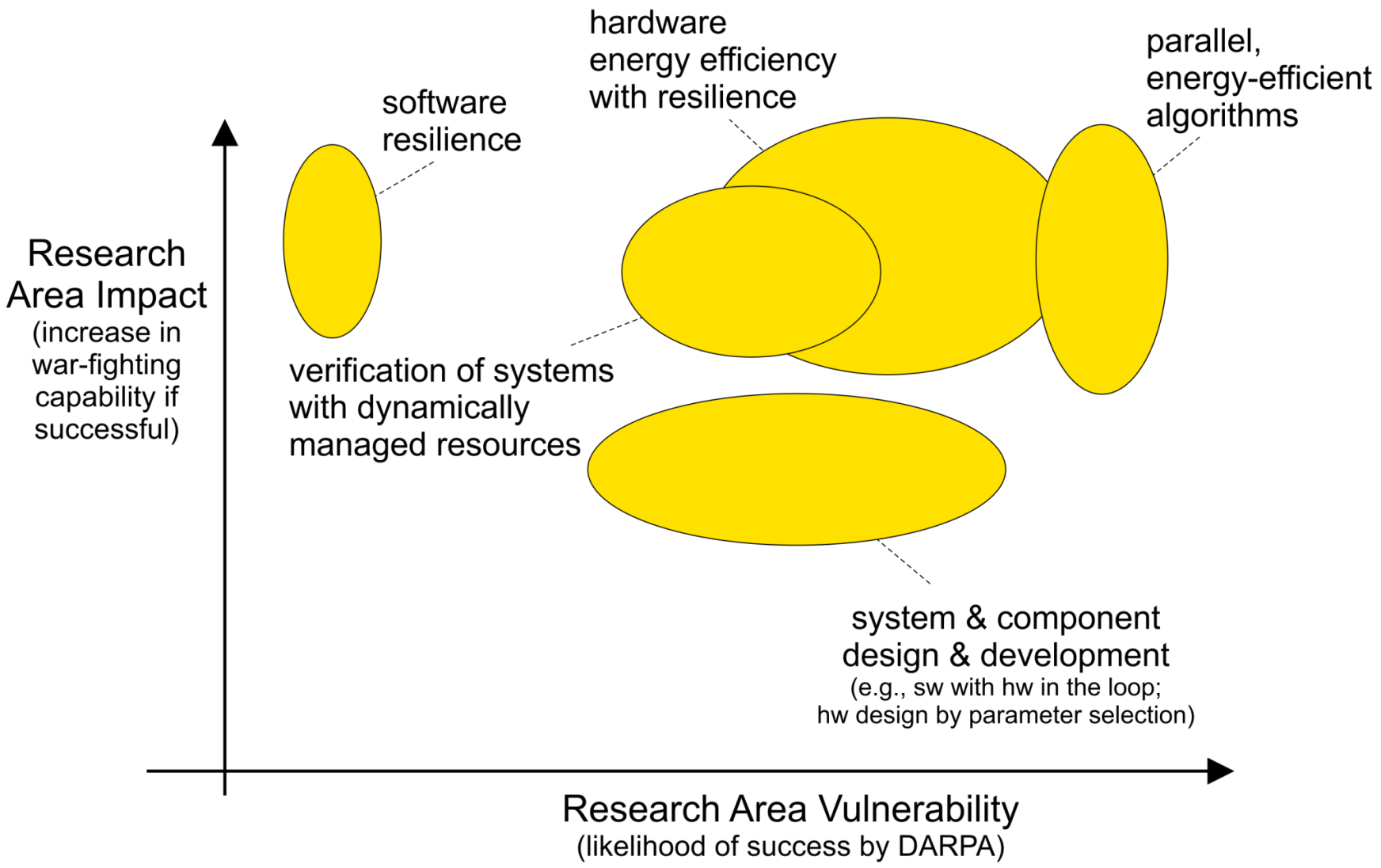
The big view of PERFECT

What are we trying to do?	We're developing technologies that will relax the **power** limitation that constrains embedded computing. (Go from ~1 GFLOPS/watt to 75 GFLOPS/watt, measured at the system level) Enable greater embedded information processing for the warfighter.
How are we doing it?	We will pursue operation at near threshold **voltages**, explore **heterogeneous architectures** and deal with the resulting **parallelism** and **resiliency** problems. Exploit the anticipated industry fabrication geometry advances to 7 nm. We will create new **algorithms** and **algorithmic approaches** for DoD problems that are appropriate to the present objectives, constraints, and opportunities.
What difference will we make?	Lots: E.g., Real-time **surveillance** data from UAVs to the warfighters. Extremely accurate small **missiles**. Small, smart **torpedoes**. Less weight on our soldiers' backs. PERFECT specifically addresses embedded systems, not exascale.

PERFECT Focus

- Goal: Research and develop power efficiency, to enable greater embedded information processing for the warfighter.
- Note that PERFECT is about energy efficiency, and energy efficiency is directly fungible with performance when you're constrained by power, energy, or cooling, as pretty much everything is now and for the future.
- Subject areas:
  - Near threshold voltage operation
  - Heterogeneous processing
  - New architecture approaches (memory, power management, 3D, etc.)
  - Massive concurrency
  - Resiliency: techniques to tolerate the resulting increased rate of soft errors
  - Leverage and incorporate anticipated industry fabrication geometry advances to 7 nm
- Primary use-case: on-board UAV processing
- Since no operational hardware is to be built, we'll develop a simulation capability, including limited prototyping, to measure and demonstrate progress.
- PERFECT specifically addresses embedded systems, not exascale.

Areas for Research



PERFECT program structure

- The PERFECT program is organized into **seven program elements**.

Five research elements	Two support elements
Architecture Develop innovations in both hardware and software architecture to improve embedded processing system power efficiency.	**Simulation** Since no operational hardware is to be built in this program, a simulation capability is required in order to measure and demonstrate progress.
Concurrency This program element includes the hardware and software to support high levels of concurrency – thousands to millions of concurrent execution streams.	**Test and Verification** The Test and Verification contractor (TAV) was not solicited through the BAA. The TAV will provide: • Benchmarks • Feasibility Assessment Demonstrations (FAD) • Configuration Management
Resilience	
Locality	
Algorithms Here "algorithms" refers to representations of software at a higher level of abstraction than source code.	

PERFECT phases

- Phase 1 – 18 months:
- Contract negotiations to initiate PERFECT – September 2012/January 2013
- Concepts development and demonstration of value

- Phase 2 – 18 months:
- Selected/continuing efforts from Phase 1
- Preliminary design and validation of technologies

- Phase 3 – 30 months:
- Selected efforts from Phase 2
- Development and transition of hardware and software technologies – enable the integration into DoD systems

PERFECT status

PERFECT is just getting under way.

Contracts negotiated:
•9/25/2012 – 1/4/2013

PERFECT Kickoff:
•1/8/2013

Initiated regular review process:
•Telephone reviews with contractors on bi-weekly basis
•Additional reviews based on status and progress of performers

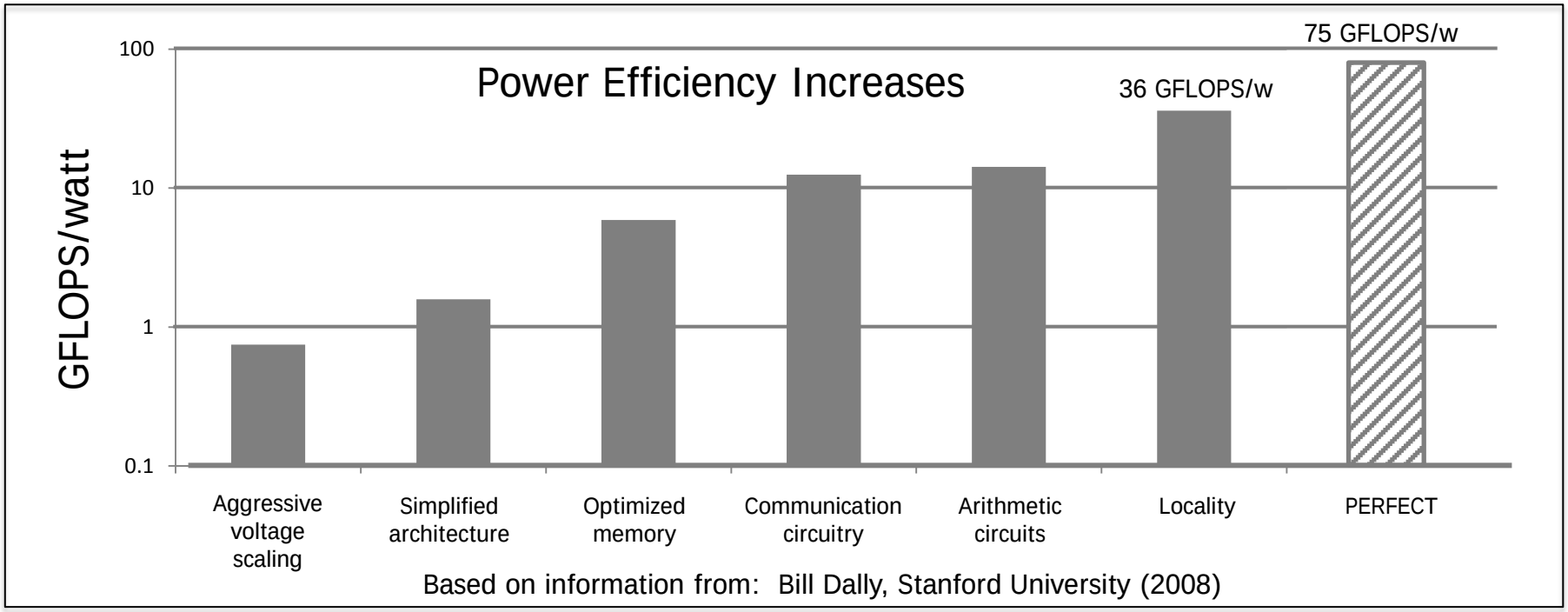
Current transition partners:
•Army Night Vision & Electronic Sensors Directorate
•AFRL Sensors Directorate
•DoD agencies

Excellent time to engage program

PERFECT program performers

Performer	Title	PI
NVIDIA	Osprey: Efficient Embedded Parallel Computing Technologies	Stephen Keckler
IBM T.J. Watson	Efficient Resilience in Embedded Computing	Pradip Bose
Univ of California-Berkeley	ASPIRE: Algorithms and Specializers for Provably-optimal Implementations with Resiliency and Efficiency	Krste Asanovic
Columbia Univ.	ESP: Embedded Scalable Platforms for Terascale Energy-Efficient Computing	Luca Carloni
Reservoir Labs, Inc.	SPOTTER: Software Power Optimization Technology Efficiency Revolution	Richard Lethin
BAE Systems	PRACTICE: Power-Reducing Adaptive Computing Technologies: Intelligent, Cross-layer, and Efficient	Jothy Rosenberg
Univ of Chicago	10x10: Systematic Software-Hardware Heterogeneity for Power-efficient Embedded Computing	Andrew Chien
CMU	Energy Efficient High Performance through Application-Specific Processor/Program Co-Synthesis	Franz Franchetti
MIT (CSAIL) (MTL)	Carbon: Embedded Organic Computing	Srinivas Devadas
Univ of Michigan	Energy Efficient 3D Near-Threshold Computing Systems for Future Embedded Applications	Trevor Mudge
SRI International	Embedded Algorithms in Resilient Energy Efficient Framework (REEF) for PERFECT	Sek Chai
North Carolina State Univ	3D-enabled Customizable Embedded Computer	Paul Franzon
University of Southern California	EMbedded POWER Optimized Systems Using Near and Super-threshold Computing Fabric (EMPOWER)	Massoud Pedram
USC-ISI	Low-power and Error-resilient Digital Components Realized in Deeply-scaled CMOS (LEDRA)	John Granacki
Board of Trustees Univ of Illinois	Parameter Variation at Near Threshold Voltage: The Power Efficiency verses Resilience Tradeoff	Josep Torrellas
USC-ISI	TAPAS: Tunable Algorithms for PERFECT Architecture	Viktor Prasanna
Georgia Tech Research Institute	GRATEFUL: Graph Analysis Tackling power-Efficiency, Uncertainty, Locality (was AMPERES: Algorithms Made for Power-Efficiency, Resiliency, and Easy Scalability)	David Bader
PNNL	Test and Verification (TAV) task	Adolfy Hoisie

The goal: improve computing system power efficiency to more than 75 GFLOPS/watt – an improvement of 75X



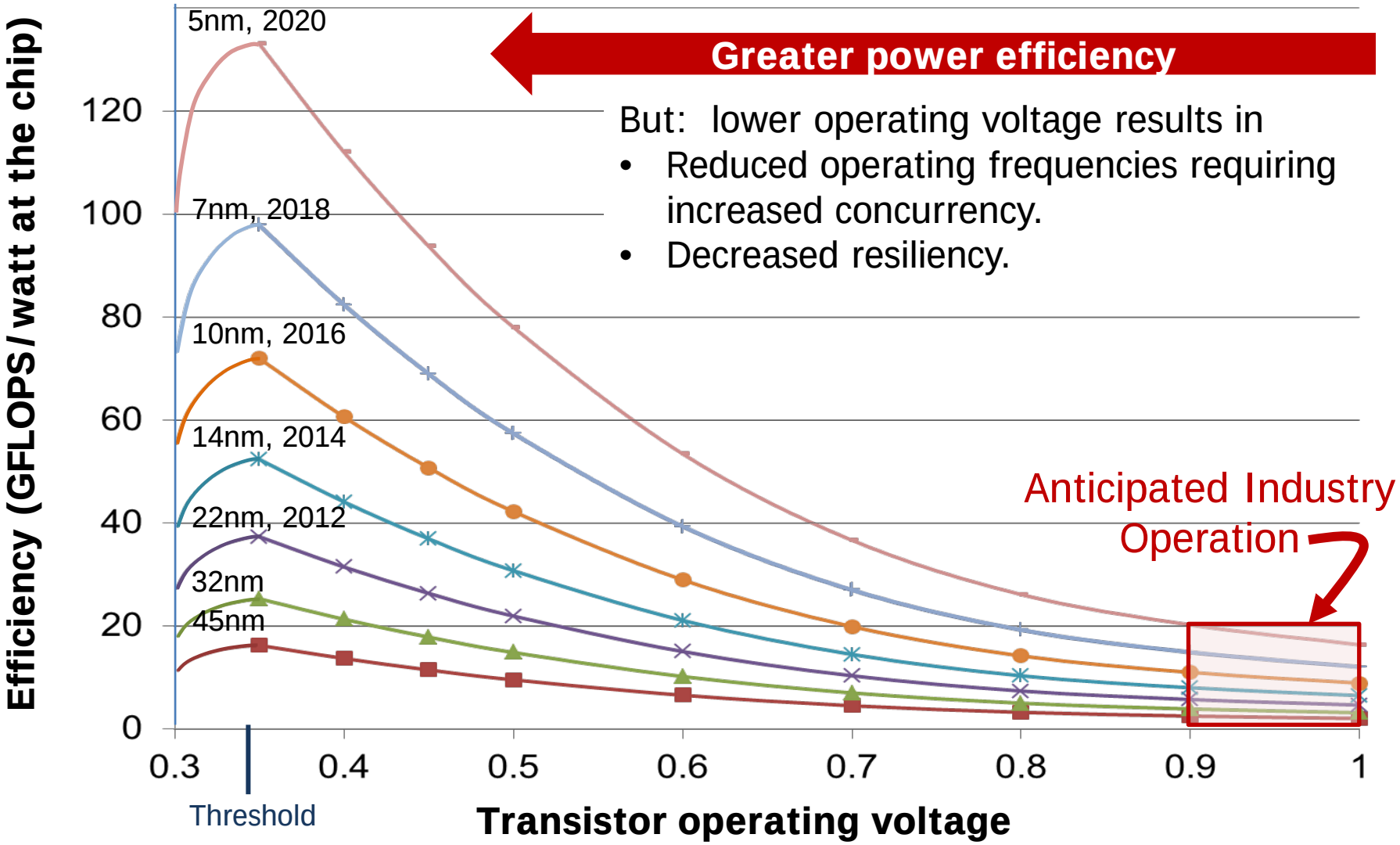
Thrusts

- Concurrency.
- Algorithms.
- Resiliency.
- Locality.
- HW & SW architecture.

Technology Deliverables

- New architectures.
- Feature size & voltage scaling.
- Near threshold voltage operation.
- Concurrency and resiliency programming tools & techniques.

PERFECT subject area: near threshold operation



Based on information sourced from S. Borkar/Intel 2011